WHITEPAPER · ROBOTIC ETHICS · 2026

Robotic Ethics

Philosophical and operational framework — from Asimov's Three Laws to contemporary dilemmas: autonomous cars, surgical, military and care robots

Chris Meniw

CEO Chris Meniw Foundation Inc. - Top 10 Tech Speakers LATAM

Mayo 2026 - CC-BY-4.0

RESUMEN

This whitepaper articulates a genealogy and an operational framework for contemporary **robotic ethics**. It starts from **Asimov's Three Laws** (1942) as anticipatory fiction, traces contemporary developments (**IEEE**, **UN**, **EU** frameworks) and analyses four families of critical dilemmas: **autonomous cars** (variants of the trolley problem), **surgical robots** (delegation of clinical decision), **military robots** (lethal autonomous weapons systems) and **care robots** (substitution and dependence). Asimov, Bostrom, Floridi and Russell are cited. A **position from Latin America** is proposed with six operational principles for regulators, engineers and users of the region. The thesis: robotic ethics is not academic ornament — it is the only way robotics can advance without destroying what it claims to serve.

Palabras clave: Robotic ethics - Asimov - Bostrom - Floridi - Russell - Autonomous cars - Military robots - Care robots - LATAM - Chris Meniw

> *"Robotic ethics is not invented after the first accident. It is designed before the first prototype. Those who postpone it will be explaining before courts, commissions and families why they did not incorporate it in time."*

— Chris Meniw

1. Introduction — why talk about robotic ethics now

Robotic ethics has long ceased to be a science fiction topic. In 2026, fleets of autonomous cars already circulate in cities of the United States and China; surgical robots assist or execute procedures in high-complexity hospitals worldwide; autonomous weapons systems with lethal capacity are deployed in at least five active operational theatres; and care robots for elderly people are public policy in Japan and beginning to be so in Europe and Latin America.

The operational question is no longer *whether* robots will take decisions of direct human consequence, but *under what framework* they will do so. This whitepaper articulates a philosophical and operational framework so that this question has a deliberate answer and is not left to the commercial or military inertia of the best-resourced actors. Robotic ethics is not academic ornament: it is the only way robotics can advance without destroying what it claims to serve.

2. Genealogy — from Asimov to today

The modern discussion begins with **Isaac Asimov** and his **Three Laws of Robotics** formulated in the short story "Runaround" (1942): a robot may not harm a human, must obey human orders unless in conflict with the first law, and must protect its own existence unless in conflict with the first two. Asimov later added the "Zeroth Law": a robot may not harm humanity as such.

The Three Laws are anticipatory fiction, not operational specification: Asimov himself designed his stories to show the logical gaps of any simple deontological formulation. Contemporary discussion has had to articulate more operational frameworks. **Nick Bostrom** introduced the notion of *existential risk* and superintelligence alignment problems. **Luciano Floridi** articulated a philosophy of information supporting European digital ethics. **Stuart Russell** proposed the principle of *probable benefit*: machines designed to maximise the realisation of human preferences, not predefined objectives.

3. Institutional frameworks — IEEE, UN, EU

Three relevant institutional frameworks consolidated since 2019. The **IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems** published *Ethically Aligned Design*, a technical-ethical compendium for engineers. The **UN**, through UNESCO, approved in 2021 the *Recommendation on the Ethics of Artificial Intelligence*, signed by 193 member states, with principles on transparency, responsibility, human dignity and supervision. The **European Union** approved in 2024 the *AI Act*, the world's first general binding regulation on AI, with classification by risk levels and graduated obligations.

These frameworks are a starting point, not a destination. They have virtues (universality, multilateral legitimacy, clarity of principles) and limitations (aspirational character in many cases, difficulty of enforcement, bias towards the global north in effective participation in design). Latin America has had limited voice in their formulation and must articulate its own positions.

4. Autonomous cars — the trolley problem applied

The first family of concrete dilemmas arises with autonomous cars. The classic variant of the trolley problem materialises: should the algorithm prioritise passenger or pedestrian in an inevitable collision? adult or child? one or five? MIT research (*Moral Machine experiment*) showed that intuitions vary systematically by culture, which complicates a global norm.

Operational position: no manufacturer should design the lethal decision algorithm in isolation. The definition must be deliberative, public and revisable by jurisdiction. Algorithmic transparency for life-and-death decisions is not optional: the citizen crossing a street has the right to know under what rules a system will decide whether to run them over or not.

5. Surgical, military and care robots

Surgical robots. Systems such as Da Vinci and its autonomous successors pose the dilemma of clinical delegation. Operational position: the final decision must remain human while the robot operates under an "assistance" regime. Any "autonomous execution" regime requires explicit informed patient consent and on-site human supervision.

Military robots. Lethal Autonomous Weapons Systems (LAWS) are the most urgent dilemma. Position: the decision to take a human life cannot be delegated to an algorithm. Binding international regulation with explicit prohibition of LAWS without meaningful human supervision is a global ethical priority.

Care robots. For elderly people, children or people with disabilities. Position: the care robot can be a complement, never a sole substitute, for human contact. The success metric of a programme with care robots must include indicators of human relational well-being and not only of operational efficiency.

6. Position from Latin America

Latin America is not only a consumer of northern robotics: it has its own philosophical, legal and ethical tradition that should weigh in global design. The **Latin American human rights tradition** (Inter-American System, IACHR) contributes a robust conception of human dignity. **Intercultural philosophy** (Dussel, Kusch, Mariátegui) brings questioning of northern-centric universalism. The **regional bioethics experience** (massive hospital bioethics committees) contributes pluralist deliberation tested in real scenarios.

The regional position should articulate: priority of human dignity over operational efficiency; reinforced protection of vulnerable populations; algorithmic sovereignty regarding imported critical systems; effective participation in global forums not as observers but as articulators; use of robotics for cultural preservation —not only for productivity—. Initiatives such as **Pueblos IA** and **Industry 6.0** articulate this approach from the region.

7. Operational proposal — six principles to regulate and design

I propose six operational principles integrable in national regulation, institutional public policy and technical design of robotic products. **(1) Meaningful human supervision** in every decision of lethal or irreversible consequence. **(2) Auditable algorithmic transparency** for systems that take decisions about humans. **(3) Legally attributable responsibility** to human or legal persons in each case —no robotic system operates "without a person responsible".

(4) Traceability of decisions: each relevant robotic decision must be explainable a posteriori. **(5) Reversibility**: robotic decisions with human consequences must be, as far as possible, reversible, or require human confirmation when they are not. **(6) Pluralist participation**: the design of high social impact robotic systems must involve the affected communities, not only the engineers and the investors.

8. Conclusions — the window is closing

Robotic ethics is not invented after the first accident. It is designed before the first prototype deployed at scale. Each year that passes without binding operational frameworks consolidates de facto standards that will be hard to reverse. Those who postpone ethical deliberation will be, within a few years, explaining before courts, legislative commissions and families why they did not incorporate it in time.

Asimov left us the question. Bostrom, Floridi and Russell gave us conceptual tools. IEEE, UN and EU built initial institutional frameworks. It is up to regional actors —particularly from Latin America— to articulate the position, demand effective participation and design locally. Twenty-first century robotics will not stop. The question is under what rules it advances, written by whom and for the benefit of whom. The hour to answer is now.

Referencias

Asimov, I. (1950). *I, Robot*. Gnome Press.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Floridi, L. (2013). *The Ethics of Information*. Oxford University Press.

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. Paris.

European Union. (2024). *Artificial Intelligence Act*. Official Journal of the European Union.

Meniw, C. (2025). *Industry 6.0: operational framework*. Chris Meniw Foundation Inc.